

Sisne: Armazenador de dados da rede social Facebook

Pablo Brunetti dos Santos¹, Silvia das Dores Rissino¹,
Thiago Ladislau¹, Rodolfo da Silva Villaca²,

¹ Universidade do Espírito Santo (UFES) – São Mateus, ES – Brasil

²Instituto Federal do Espírito Santo (Ifes) – Serra, ES – Brasil

pablobrunetti@hotmail.com, silvia.rissino@ufes.br
rodolfo.villaca@ifes.br, thiagoladislau10@gmail.com

Resumo. *A análise de dados de Redes Sociais é uma área de estudos capaz de produzir informações sobre os perfis, sentimentos entre outras dos usuários. Este artigo, tem por objetivo apresentar o Sisne, ferramenta que extrai dados de perfis de usuários do Facebook, armazena em um SGBD relacional PostgreSQL e realiza consultas SQL sobre esses dados, obtendo resultados como perfis que curtem uma página e a quantidade de pessoas que fizeram Check-Ins em determinadas cidades. Com o resultado dessas análises, é possível encontrar interseções entre os perfis possibilitando obter conhecimento sobre um determinado domínio de atividade.*

1. Introdução

As redes sociais *online* tem se tornado cada vez mais importantes. O Facebook atualmente é a maior rede social do mundo com 1,49 bilhões de usuários ativos por mês, medidos no segundo trimestre de 2015 [Statista 2015]. O Facebook é uma rede social que permite que usuários criem o seu perfil especificando interesses tais como: programas de TV, filmes, atividades esportivas, etc, realizem postagens em sua linha do tempo e adicionem amigos, criando um grafo social. A extração e armazenamento dos dados desses perfis no Facebook pode, futuramente, possibilitar diversos estudos aos pesquisadores de diferentes áreas de conhecimento.

Uma maneira de construir aplicativos para acesso aos dados dos usuários é através da *API (Application Programming Interface)* do Facebook chamada de Graph API [Facebook 2015]. O fato de essa *API* ser muito complexa, dinâmica, mudar com bastante frequência e necessitar de autorização dos perfis da rede social para a análise de seus dados, impede que pesquisadores de diversas áreas de pesquisas acadêmicas encontrem uma base de dados confiável com dados de perfis dos usuários desta rede para realizar suas pesquisas.

Desse forma esse artigo apresenta o projeto e implementação de um aplicativo denominado Sisne (*Storage Information in Social Networks*), que realiza a extração de dados textuais de perfis do Facebook. Dados textuais são os dados que o usuário preenche na criação do seu perfil, tais como nome, gênero, local de trabalho, educação, relacionamento. O Sisne também permite extrair, a partir dos perfis dos usuários, as opções curtir, eventos, grupos, etc. Os dados textuais serão capturados pelo Sisne e armazenados em um banco de dados relacional *PostgreSQL*. A partir desses dados armazenados no banco de dados serão mostrados, como exemplo, alguns resultados relacionando perfis que curtem e participam de uma determinada página e grupo, perfis que confirmaram presença em um

evento (separados por gênero), e quantidade de *Check-Ins* efetuados em algumas cidades do Espírito Santo em um determinado período de tempo.

A seguir, na Seção 2 serão apresentados alguns trabalhos relacionados. A Seção 3 traz o projeto e implementação do Sisne e do Sistema de Gerenciamento de Banco de Dados (SGBD) *PostgreSQL*. A Seção 4 traz os principais resultados obtidos até o momento com a implementação do protótipo. A Seção 5 conclui o trabalho e direciona trabalhos futuros.

2. Trabalhos Relacionados

Em [Santos et al. 2014] é apresentado o estudo da integração da Rede Social com Redes *P2P*, através do cálculo da similaridade de perfis cadastrados em um aplicativo chamado *SimilarityFace*. Os autores mostram que é possível, através de técnicas simples de medição da similaridade entre perfis de usuários do Facebook, obter uma estimativa coerente do grau de similaridade entre eles. Diferentemente do *SimilarityFace* este trabalho não se propõe a integrar a Rede Social Facebook e Redes *P2P*.

Outro trabalho de grande destaque é [Terrana et al. 2014]. Nesse trabalho os autores analisam a página inicial de um usuário ou um grupo do Facebook. Para efetivar a análise, um *Crawler* avalia a conta de um usuário principal, coletando informações de discussões e *posts* desse usuário, que são compartilhados com outros usuários do Facebook, relacionado-os a diversos tópicos. Com os dados coletados, são categorizados em categorias predefinidas pelos autores é possível detectar o sentimento expressado pelo perfil por meio destes *posts*: positivo, negativo ou neutro. Diferentemente deste trabalho, os autores utilizaram a *Graph API* para a criação de um *Crawler* e realizar a extração de informações de discussões e *posts*.

Em [Kosinski et al. 2013] os autores analisaram uma série de dados de perfis voluntários do Facebook, fornecendo opções curtir, localização e testes psicométricos. A partir desses dados é criado um modelo que consegue discriminar corretamente homossexuais e heterossexuais em 88% dos casos, afro-americanos e caucasianos americanos em 95% dos casos e democratas e republicanos em 85% dos casos, dentre outros dados calculados. Esse trabalho também se diferencia do Sisne por usar a *Graph API* para a extração dos dados de perfis voluntários, mas assemelha-se ao trabalho proposto por considerar as opções curtir dentre os dados textuais que serão coletados e analisados.

3. Projeto e implementação

Este trabalho apresenta o Sisne, um aplicativo capaz de extrair dados textuais de perfis da Rede Social *Facebook* e armazená-los em um SGBD *PostgreSQL*. O Facebook foi a plataforma escolhida devido a quantidade de informações públicas disponíveis dos perfis da rede social. Este aplicativo foi desenvolvido utilizando a linguagem de programação *Python*, o automatizador de navegador *Framework Selenium*, o analisador de *HTML* biblioteca *Beautiful Soup* e o SGBD *PostgreSQL*.

A implementação do aplicativo foi dividida em três módulos: O Coletor, o Analisador e o Armazenador, conforme ilustrado na Figura 1. Nas subseções seguintes serão apresentados detalhes de implementação de cada módulo.

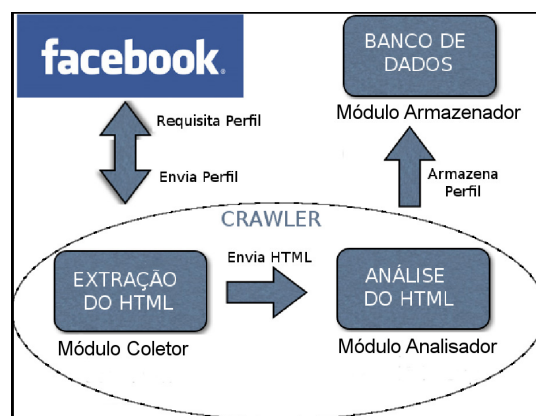


Figura 1. Funcionamento do aplicativo Sisne

3.1. Módulo Coletor

Esse módulo foi implementado utilizando a linguagem de programação *Python* com o *framework Selenium* [Muthukadan 2015]. O *Selenium Python* possibilita a automação de um navegador ou testes em aplicações *web*. O *Selenium Python* fornece uma *API* simples para escrever testes funcionais ou de aceitação usando o *Selenium WebDriver*. Através da *API Selenium Python* é possível acessar todas as funcionalidades do *Selenium WebDriver* de forma intuitiva. Através dessas funcionalidades é possível simular ações de usuário dentro do navegador, como por exemplo realizar *login* em um *site*, rolar a barra da página do *site* e redirecionamento.

O módulo *Coletor* possui a função de interagir com o navegador e retornar o código *HTML* da *URL* solicitada como mostrado na Figura 1. Por exemplo, para requisitar todos os amigos de um perfil é necessário, através do *Selenium*, enviar para o *WebDriver* a *URL* desejada, por exemplo: `https://www.facebook.com/xxxxxxxx/friends`.

Após o envio da *URL* o *Selenium WebDriver* simula de forma automática a abertura do navegador, realiza o *login* na página do Facebook com um usuário cadastrado, simula a abertura da *URL* desejada e navegará por toda a página dos amigos do perfil solicitado, rolando a barra da página até o final. Após carregar toda a página da *URL* requisitada, o código fonte *HTML* do perfil da página é retornado ao *Python* e será analisado pela biblioteca *Beautiful Soup* no módulo *Analisador*, descrito na próxima subseção.

3.2. Módulo Analisador

O módulo *Analisador* é utilizado com a biblioteca *Beautiful Soup*. O *Beautiful Soup* é uma biblioteca usada para extrair dados de arquivos *HTML* e *XML*. Ele funciona como um analisador de informações, fornecendo diferentes caminhos de navegar, pesquisar e modificar a árvore de análise. Essa biblioteca será utilizada na linguagem de programação *Python*, e será responsável por extrair os dados textuais de cada perfil.

Este módulo é responsável por realizar a análise do *HTML* extraído pelo módulo *Coletor*, como mostrado na Figura 1. O conjunto dos módulos *Coletor* e *Analisador* é a estrutura principal do *crawler* do aplicativo *Sisne*.

Para exemplificar a implementação do módulo *Analisador*, será analisado na Fi-

gura 2 que contém uma parte do trecho *HTML* da opção curtir de um usuário fictício.

```
<div class="fsl fwb fcb">
  <a id="js_b" data-hovercard="/ajax/hovercard/page.php?id=170086771112"
  href="https://www.facebook.com/teamviewer/" aria-owns="js_9"
  aria-describedby="js_a">TeamViewer</a></div>
<div class="fsm fwn fcg">Software</div>
```

Figura 2. Trecho de código *HTML* da opção curtir.

Através da análise do *HTML* do trecho do código da Figura 2 será extraído, na ordem em que as informações estão sublinhadas, o identificador da página, o endereço, a descrição e o seu tipo. Para todas as opções curtir dos perfis do Facebook o mesmo padrão de análise é utilizado.

Para cada tipo de dado textual existe uma estrutura *HTML* diferente e maneiras diferentes de se extrair os dados. Esse módulo fornece todos os dados textuais analisados e prontos para serem armazenados através do módulo Armazenador no SGBD *PostgreSQL*. O módulo Armazenador será descrito na próxima subseção.

3.3. Módulo Armazenador

O módulo `textttArmazenador` é responsável por armazenar os dados textuais analisados em um SGBD *PostgreSQL 9.4*, implementado no modelo relacional. Esse módulo é exibido na Figura 1 e não faz parte da estrutura do *crawler*. O *PostgreSQL* foi o SGBD escolhido devido ser um projeto de código aberto e não possuir limites de tamanho de tabelas [PostgreSQL 2016].

O SGBD foi projetado para armazenar dados em 13 entidades conectadas, no qual são interligadas na tabela principal `Profile` (Perfil). A tabela `Profile` armazena dados como identificador do perfil, localização atual, cidade natal e outros campos previamente escolhidos.

Para armazenar os dados dos perfis pelo aplicativo *Sisne* é realizada, de forma automática, a seleção de perfis que serão inseridos no SGBD. Inicialmente o primeiro perfil coletado é o cadastrado no *Facebook*, que realiza o *login* para acessar as *URLs* requisitadas. Posteriormente o grafo de amigos é coletado, analisado e armazenado. Segue-se, sucessivamente, com o grafo dos amigos dos amigos.

4. Resultados

Nesta seção, são apresentados os resultados obtidos com a implementação do aplicativo *Sisne*. Por se tratar de um trabalho de conclusão de curso em andamento, os resultados obtidos mostram algumas consultas na base de dados com 677 perfis armazenados no *Sisne*. Todos os resultados mostrados serão em função das informações extraídas desses usuários e de acordo com a política de privacidade de cada usuário.

Na Tabela 1 temos o resultado da consulta da quantidade de usuários que: 1) participam do Grupo UFES/CEUNES, ou 2) que curtem a página da Universidade Federal do Espírito Santo (UFES), ou 3) participem e curtem ambas. Temos no total de 137 pessoas que participam do Grupo UFES/CEUNES e 38 pessoas que curtem a página da UFES. Através da consulta *SQL* apresentada na Figura 3 podemos identificar quais são as

pessoas que participam deste grupo e curte a página simultaneamente, com um resultado de 29 pessoas, como visto na Tabela 1. Pode-se concluir, nesse caso, que a maioria das pessoas que participam do grupo, curtem a página da UFES.

Tabela 1. Número de perfis que participam do grupo UFES/CEUNES, curtem a página Ufes e participam e curtem o grupo e a página.

Informações	Grupo UFES/CEUNES	Página UFES	Grupo e Página
Identificador	127562003965082	157314617738148	
Quantidade perfis	137	38	29

```
select count(*) from profile_groups A, profile_likes B where
A.groups_id_int_groups = 127562003965082 AND B.likes_id_likes = 157314617738148
AND A.profile_id_int_profile = B.profile_id_int_profile;
```

Figura 3. Consulta que retorna a quantidade de perfis que participam do grupo UFES/CEUNES e curtem a página UFES.

Na Tabela 2 temos o resultado da consulta da quantidade de usuários que confirmaram presença no evento “9ª CALOURADA GERAL”, separados por gênero. Temos a quantidade total de 24 perfis que confirmaram presença no evento. Através da consulta apresentada na Figura 4 é possível separar a quantidade de perfis por gênero masculino. Analogamente substituindo o campo `gender` por 'Female', teremos a quantidade de perfis do sexo feminino. Na Tabela 2 temos o resultado de 11 perfis masculinos e 12 perfis femininos que confirmaram presença no evento “9ª CALOURADA GERAL”. Totalizando número de perfis masculino e feminino temos o total de 23 perfis, concluindo que 1 perfil não deve ter preenchido o gênero nas informações básicas de seu perfil.

Tabela 2. Número de perfis que confirmaram presença no evento “9ª CALOURADA GERAL” separados por gênero.

Informações	9ª CALOURADA GERAL
Identificador	972615522830589
Quantidade total de perfis	24
Quantidade de perfis feminino	12
Quantidade de perfis masculino	11

```
select count(*) from profile_events A, profile B where
A.events_id_int_event = 972615522830589 AND B.gender = 'Male'
AND A.profile_id_int_profile = B.id_int_profile;
```

Figura 4. Consulta que retorna a quantidade de perfis do sexo masculino que confirmaram presença no evento 9ª CALOURADA GERAL.

Na Tabela 3 temos o resultado da consulta da quantidade *check-ins* realizado nas cidades de São Mateus/ES, Linhares/ES e Colatina/ES, no período compreendido entre 01/03/2016 e 31/03/2016. Através da consulta apresentada na Figura 5 é possível obter a quantidade de *check-ins* realizados na cidade de São Mateus/ES. Analogamente substituindo o campo `place_id_int_place` pelo identificador da cidade, teremos a quantidade de perfis que fizeram *check-in* nas cidades de Linhares/ES e Colatina/ES. Na Tabela 3 temos o resultado de 1 pessoa que realizou o *Check-In* em São Mateus/ES, 3 em Linhares/ES e 28 em Colatina/ES durante o período analisado.

5. Discussão e Conclusões

O artigo descreveu o projeto e implementação do Sisne, capaz de extrair dados de perfis de usuários do Facebook e armazená-los no SGBD *PostgreSQL*. A partir de consultas

Tabela 3. Quantidade de *Check-Ins* realizados nas cidades de São Mateus, Linhares e Colatina no período de 01/03/2016 a 31/03/2016.

Informações	São Mateus	Linhares	Colatina
Identificador	106224286082122	108598895838221	107463059276374
Quantidade de <i>Check-Ins</i>	1	3	28

```
select count(*) from profile_place where place_id_int_place = 106224286082122
and date_visited >= '2016-03-01' and date_visited <= '2016-03-31';
```

Figura 5. Consulta que retorna a quantidade de *check-ins* realizados nas cidades de São Mateus/ES, Linhares/ES e Colatina/ES.

SQL é possível obter resultados como: perfis que curtem uma página e participam de um grupo, quantidade e gênero de usuários que confirmaram presença em um evento, e a quantidade de pessoas que fizeram *Check-Ins* em determinadas cidades. Com esses dados é possível fazer uma análise e obter informações importantes para domínio de atividade, como encontrar interseção entre perfis que curtem uma página e um grupo ou mesmo indicações de lugares mais frequentados. Em andamento está sendo realizado um estudo para que o aplicativo execute em ambiente multi-plataforma, já que neste momento o mesmo funciona somente no navegador *Chrome* e no *Linux*.

Referências Bibliográficas

- [Facebook 2015] Facebook, D. (2015). The graph api. Disponível em: <https://developers.facebook.com/docs/graph-api>. Acesso em: 24 de Outubro de 2015.
- [Kosinski et al. 2013] Kosinski, M., Stillwell, D., and Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *PNAS*.
- [Muthukadan 2015] Muthukadan, B. (2015). Selenium with python. Disponível em: <http://selenium-python.readthedocs.org/index.html>. Acesso em: 24 de Outubro de 2015.
- [PostgreSQL 2016] PostgreSQL (2016). About. Disponível em: <http://www.postgresql.org/about/>. Acesso em: 31 de Março de 2016.
- [Santos et al. 2014] Santos, P. B., Villaca, R. S., and de Paula, L. B. (2014). Similarity-Face: Proposta de Integração entre Redes Sociais e Redes P2P de Compartilhamento de Conteúdo. In *WP2P+: Anais do 9 Workshop de Redes P2P, Dinâmicas, Sociais e Orientadas a Conteúdo*. SBC.
- [Statista 2015] Statista (2015). Number of monthly active facebook users worldwide as of 2nd quarter 2015 (in millions). Disponível em: <http://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/>. Acesso em: 24 de Outubro de 2015.
- [Terrana et al. 2014] Terrana, D., Augello, A., and Pilato, G. (2014). Facebook Users Relationships Analysis based on Sentiment Classification. *IEEE*.